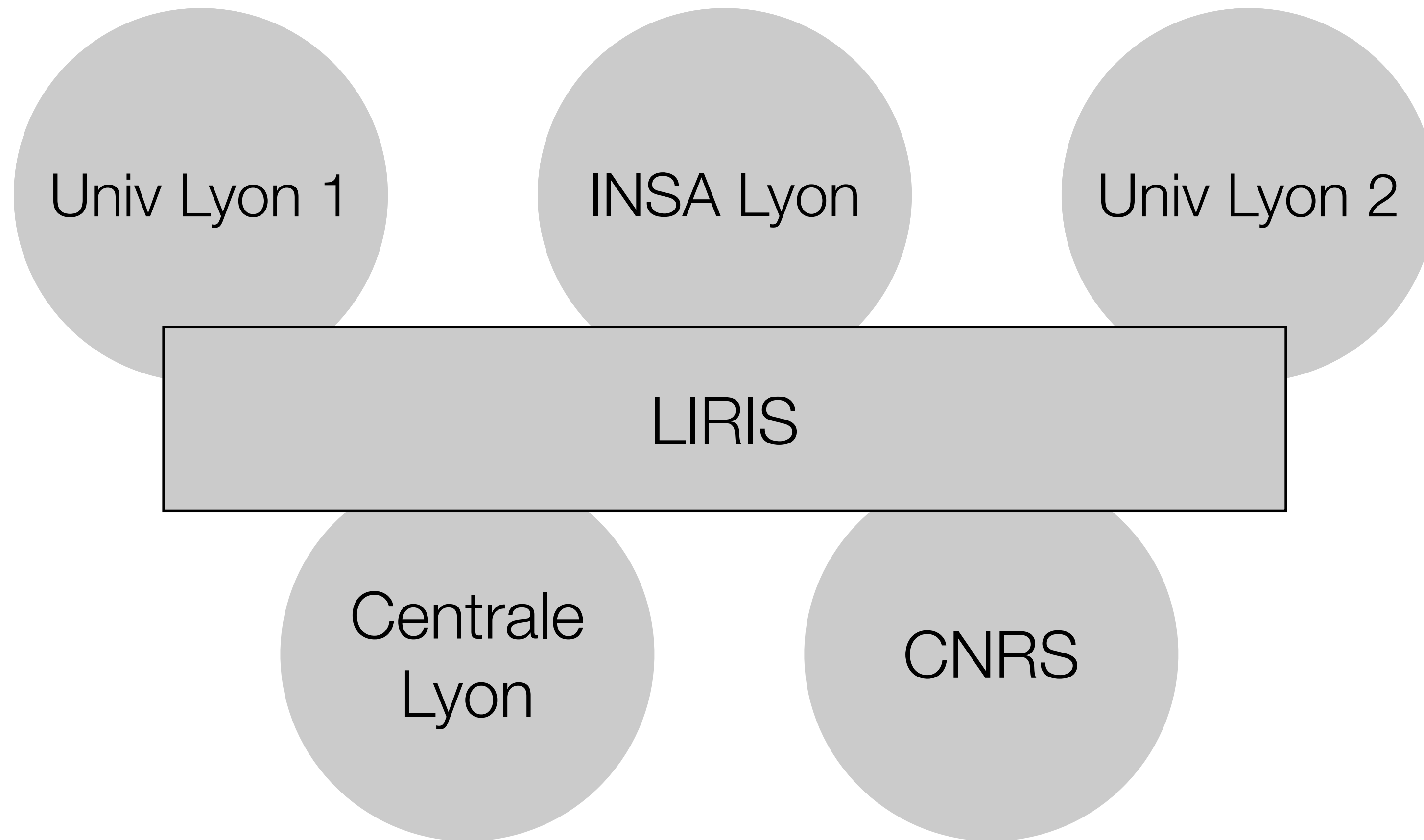


# Méthodes d'évaluation

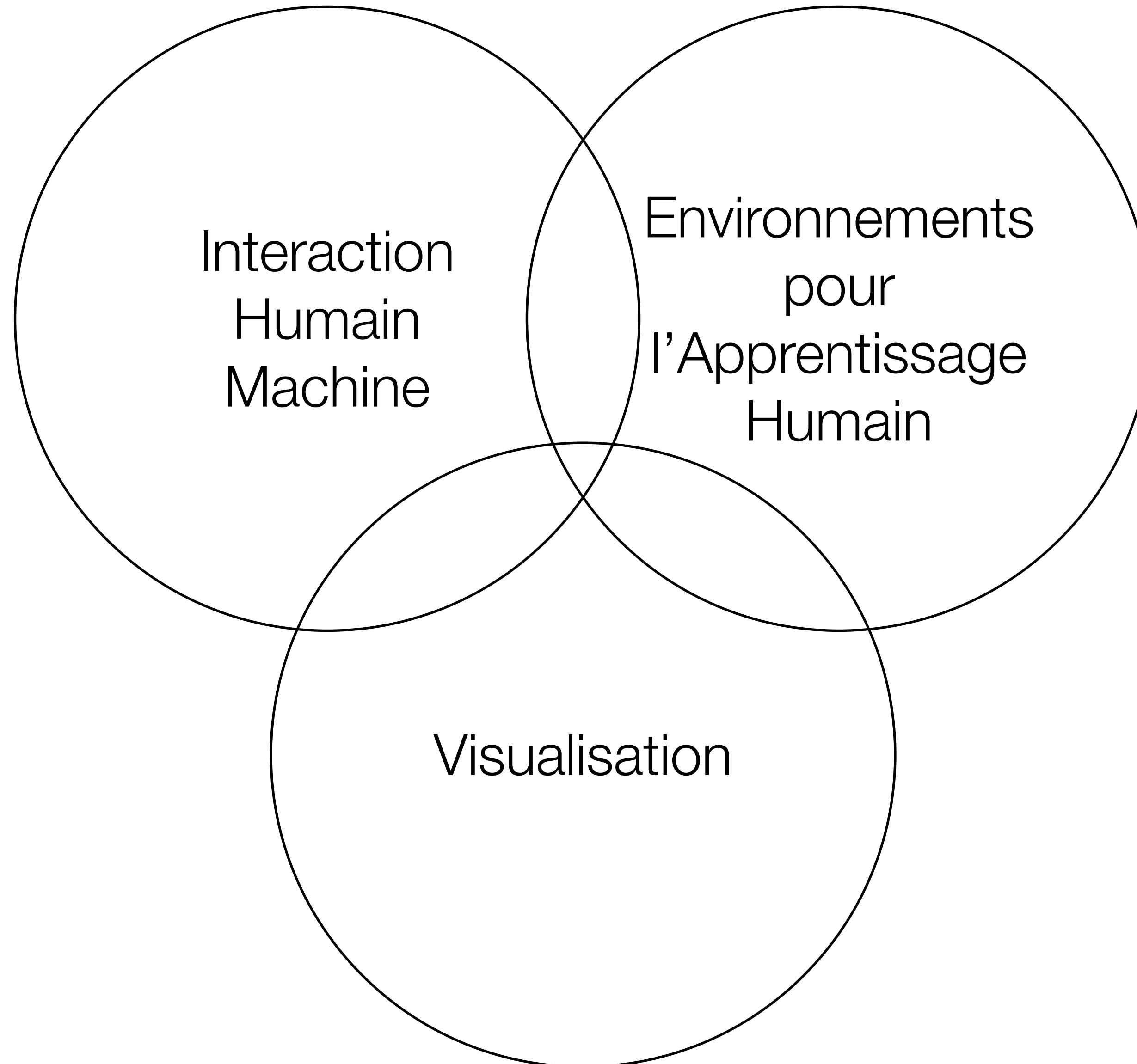
**Aurélien Tabard**

SICAL, LIRIS, Université Lyon 1

# Le LIRIS



# SICAL







**Échauffement :** Comment comparer la filtration de masques sur le risque de contamination ?





# Méthodes d'évaluation

## Matin

- Introduction ~ 50 min
  - Approches d'évaluation
  - Méthodes analytiques
  - Méthodes empiriques
- Concevoir une expérience 1h10
  - Exemples
  - Les bases
  - La structure d'une expérience
  - Mener une expérience
  - Collecter les données
- Mise en pratique 1h

# Méthodes d'évaluation

## Après midi

- Data collection + analysis setup 45 min
- Visual and statistical analysis 2h
  - Practice
  - Checking your data
  - Significance testing with t-tests
  - Significance testing with Anova
  - Measuring effect sizes
  - Beyond significance testing
- Wrap up of the analysis 15 min

# Méthodes d'évaluation

- Introduction
- **Approches d'évaluation**
- Méthodes analytiques
- Méthodes empiriques

# Analytical vs. Empirical Evaluation

*“If you want to evaluate a tool, say an axe, you might study the design of the bit, the weight distribution, the steel alloy used, the grade of hickory in the handle, etc., or you may just study the kind and speed of the cuts it makes in the hands of a good axeman.”*

Scriven, 1967



# Complementary methods

**Empirical evaluation** enables to understand the implications of the object properties

- Will the axe cut this log ?

**Analytical evaluation** offers a critical grid on important properties

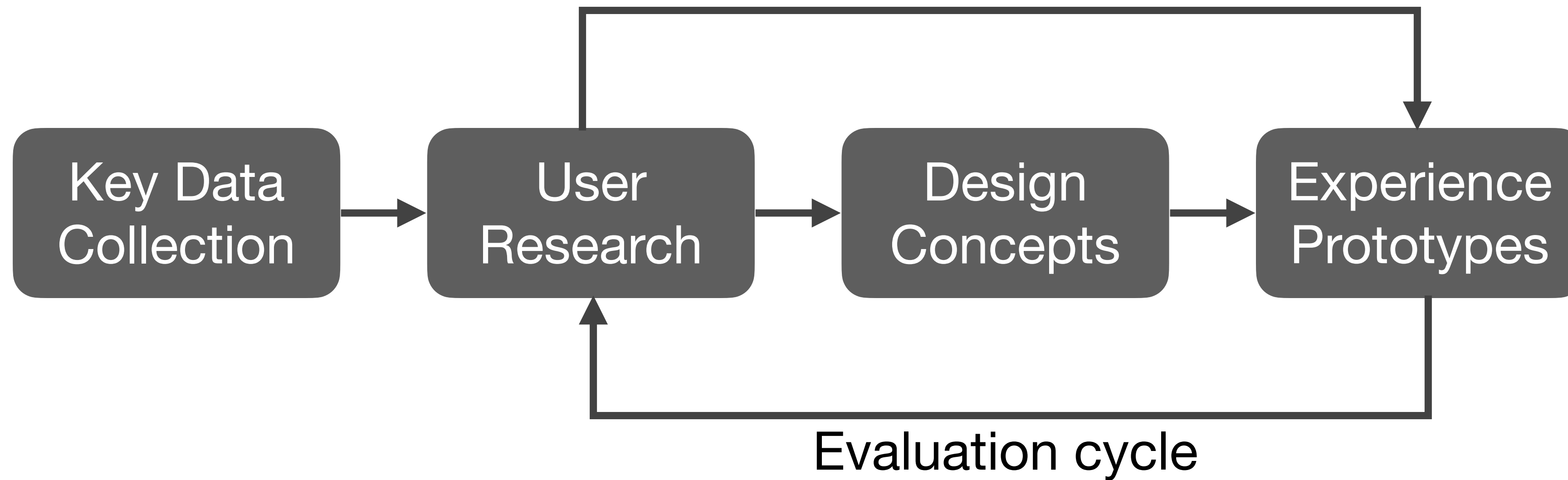
- Is the axe handle compatible with left-handed people ?

In both cases :

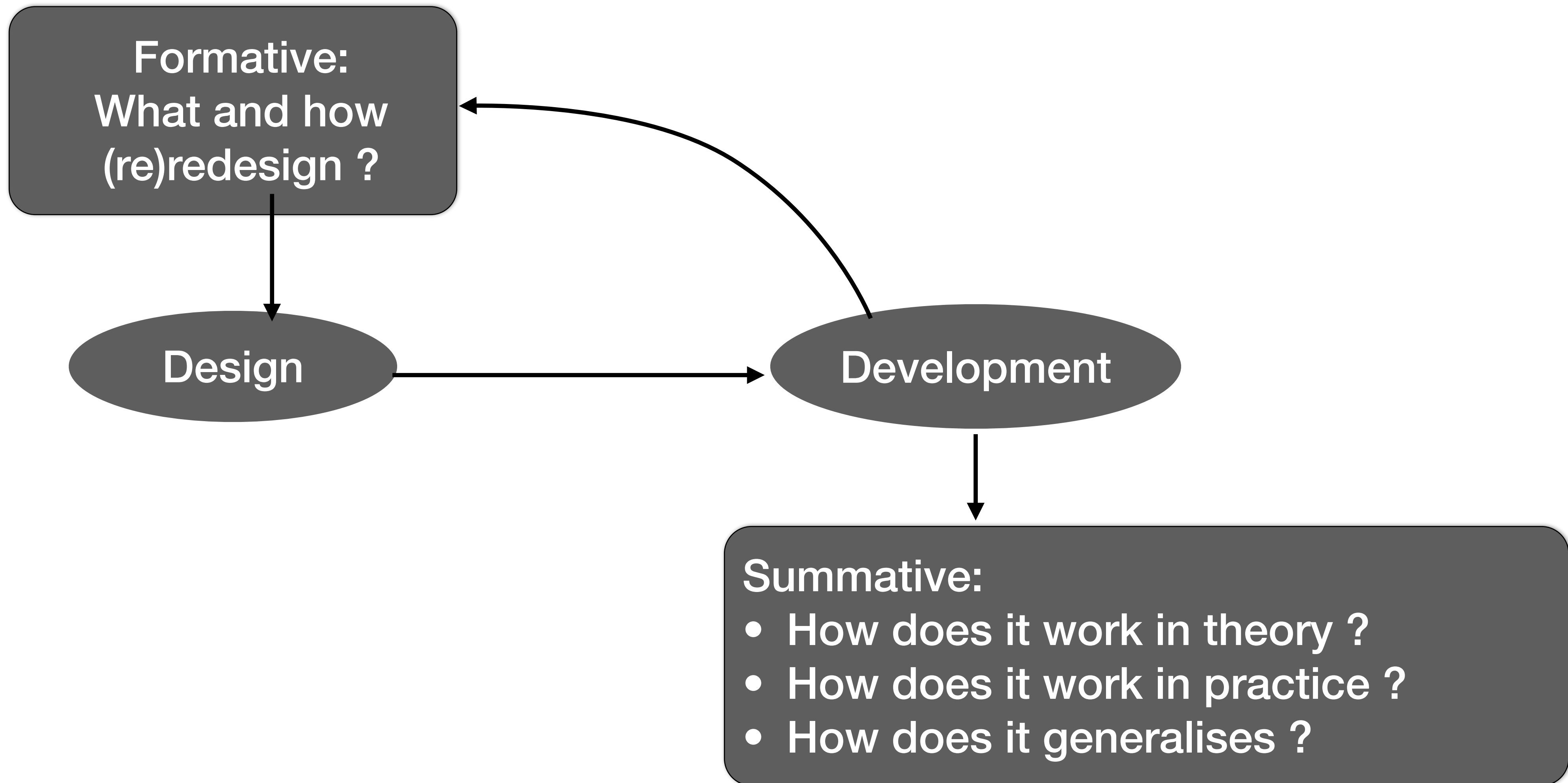
- Production of facts that must be interpreted

# Formative evaluations

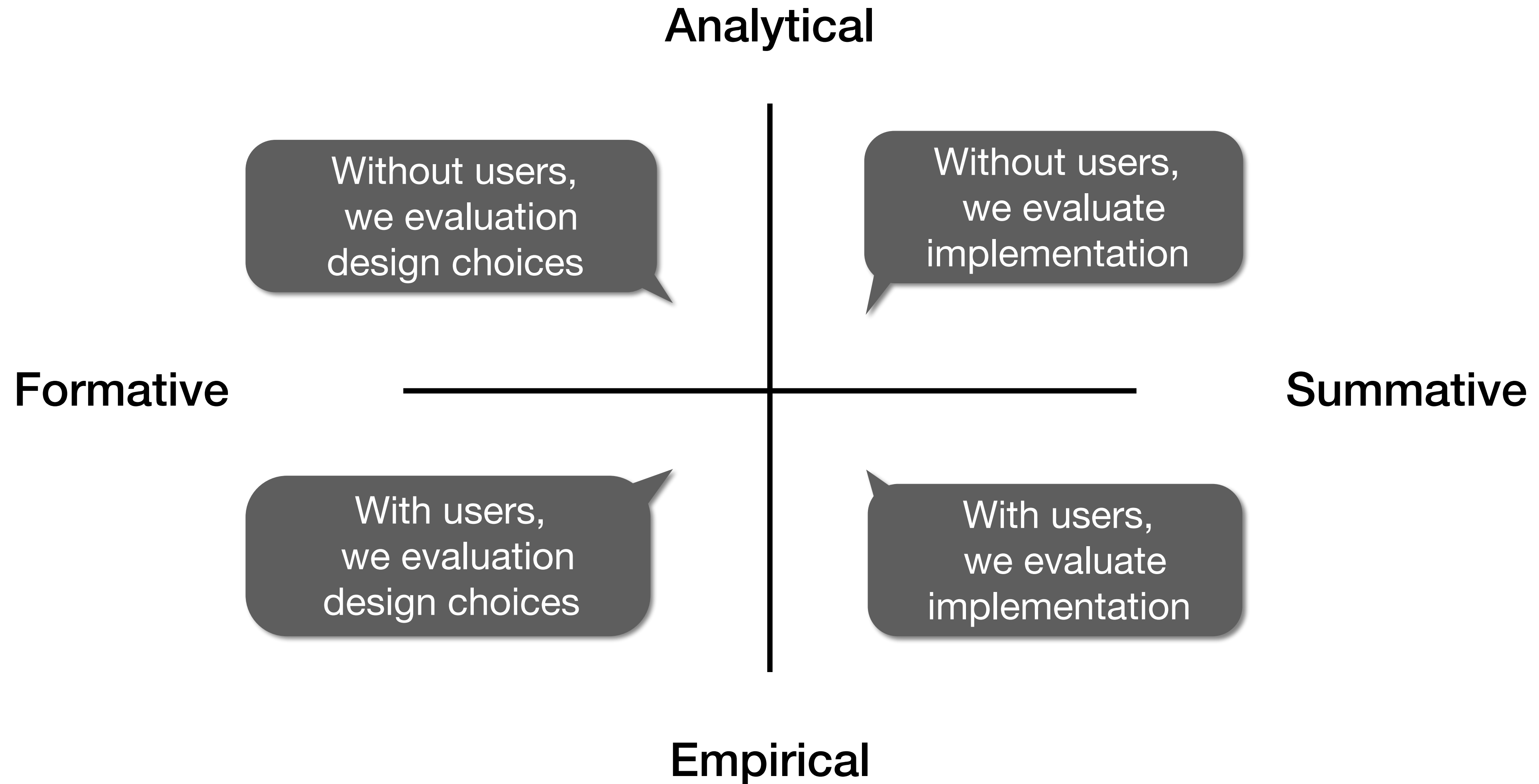
Part of a product design cycle



# Formative or Summative Evaluation ?



# Orthogonal approaches





# Evaluation without criteria is useless!

Possible criteria (among others) :

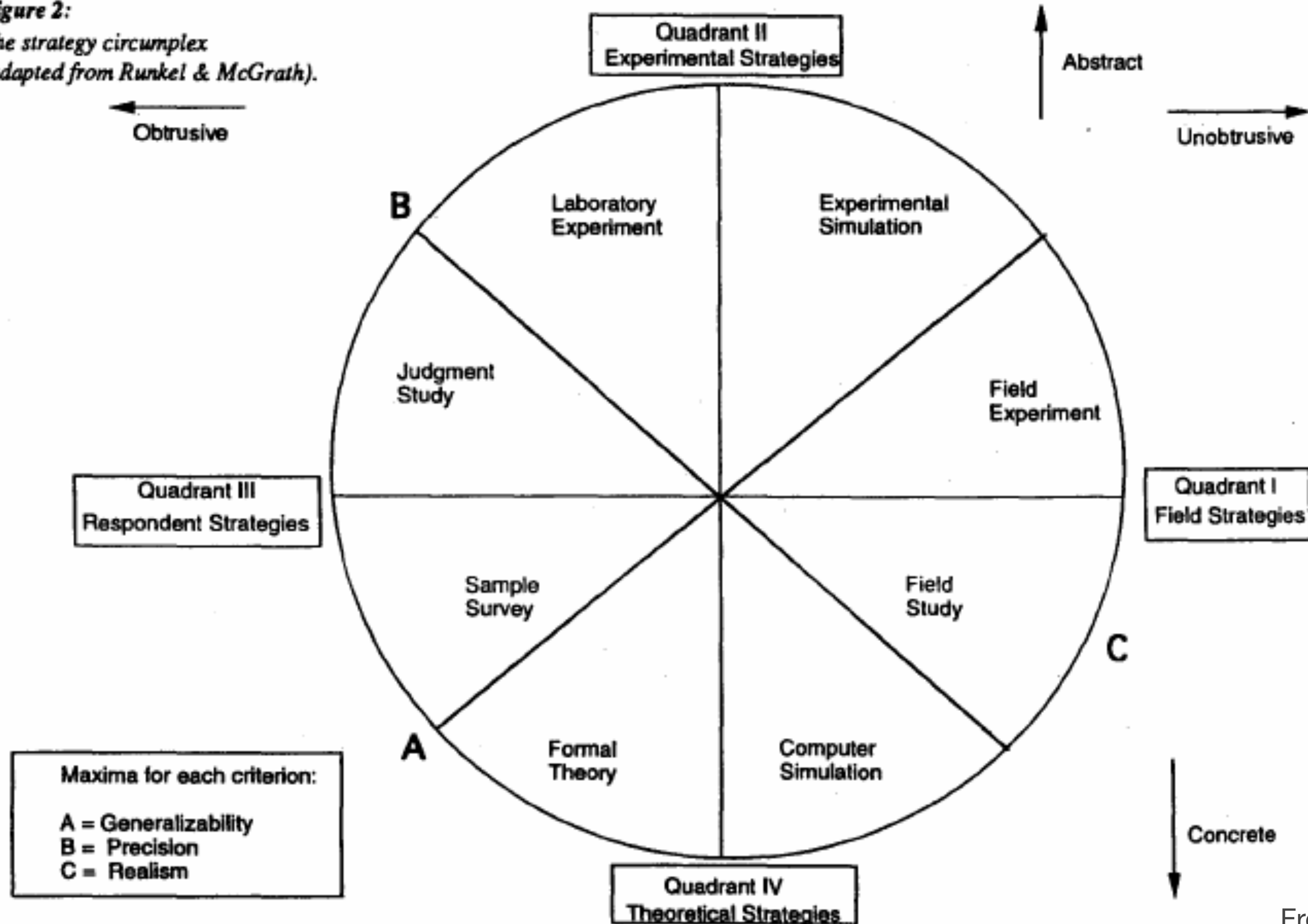
- Informal test of an idea against another
- Statistical analysis of average performance
- Acceptation by a realistic user group
- Checking heuristic / ergonomic criteria
- Usability of a given design for a given task

**Always define what you want to learn before evaluating!**

# Taxonomy of Methods

[McGrath et al. 1994]

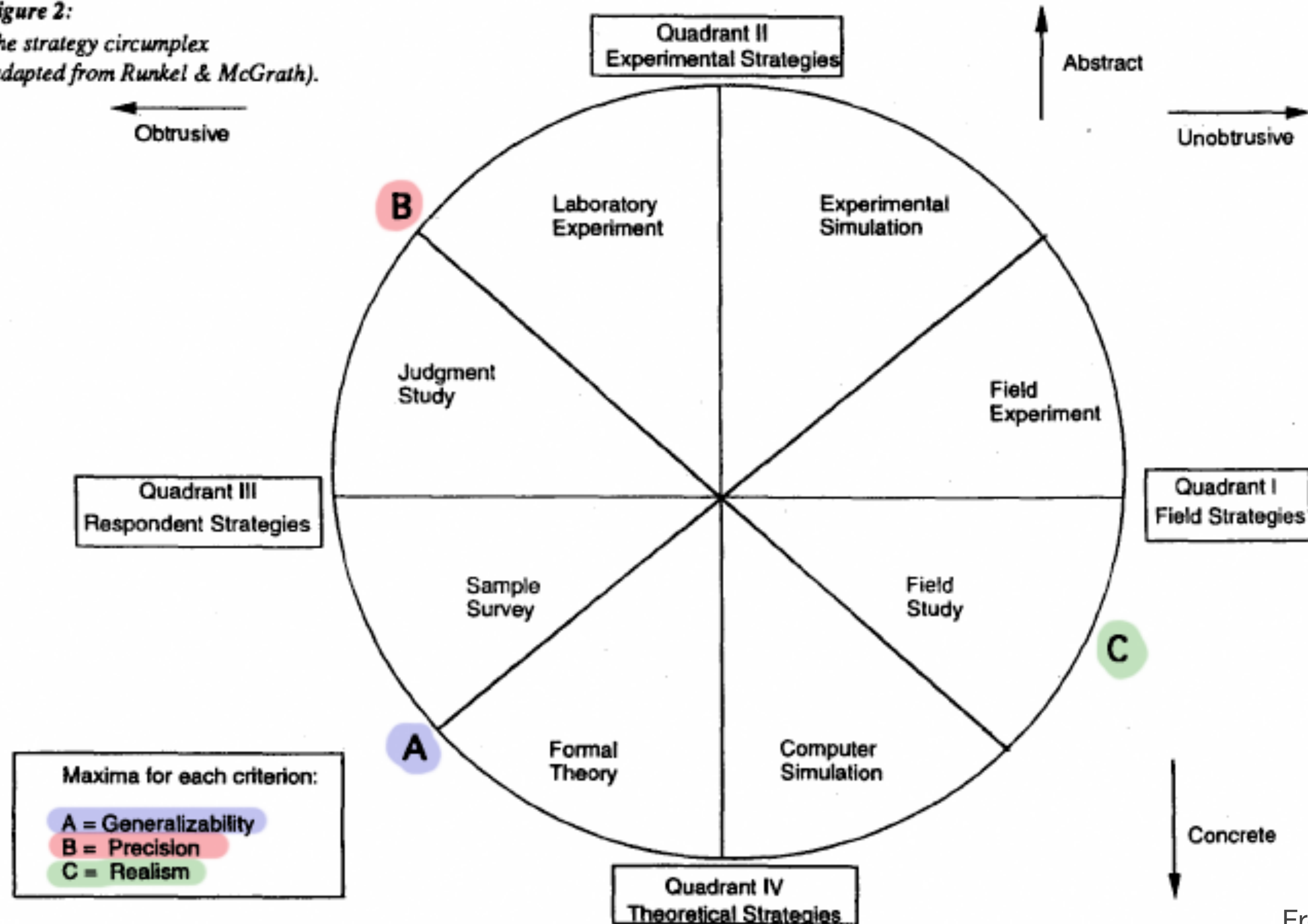
Figure 2:  
The strategy circumplex  
(adapted from Runkel & McGrath).



# Taxonomy of Methods

[McGrath et al. 1994]

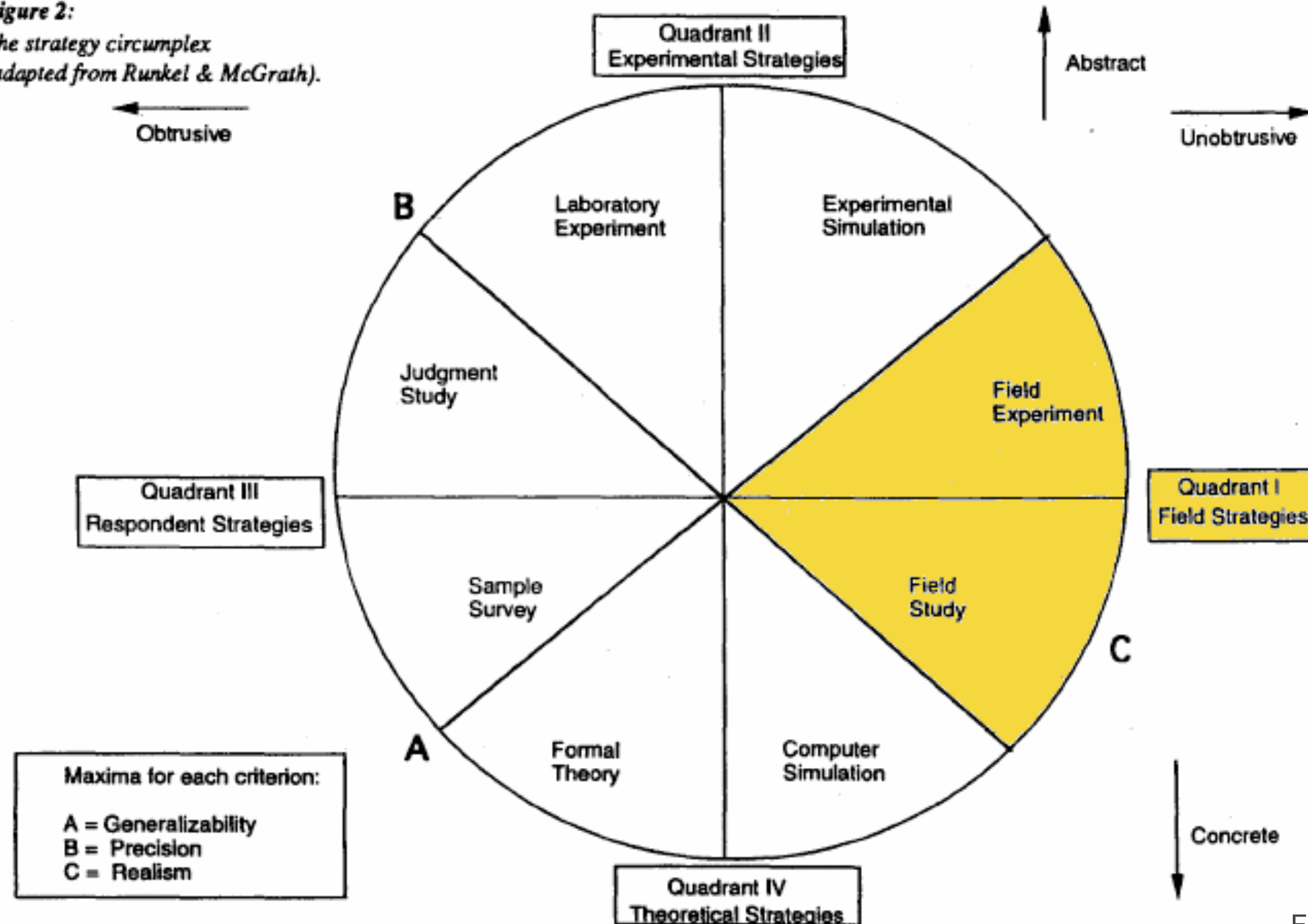
Figure 2:  
The strategy circumplex  
(adapted from Runkel & McGrath).



# Taxonomy of Methods

[McGrath et al. 1994]

Figure 2:  
The strategy circumplex  
(adapted from Runkel & McGrath).

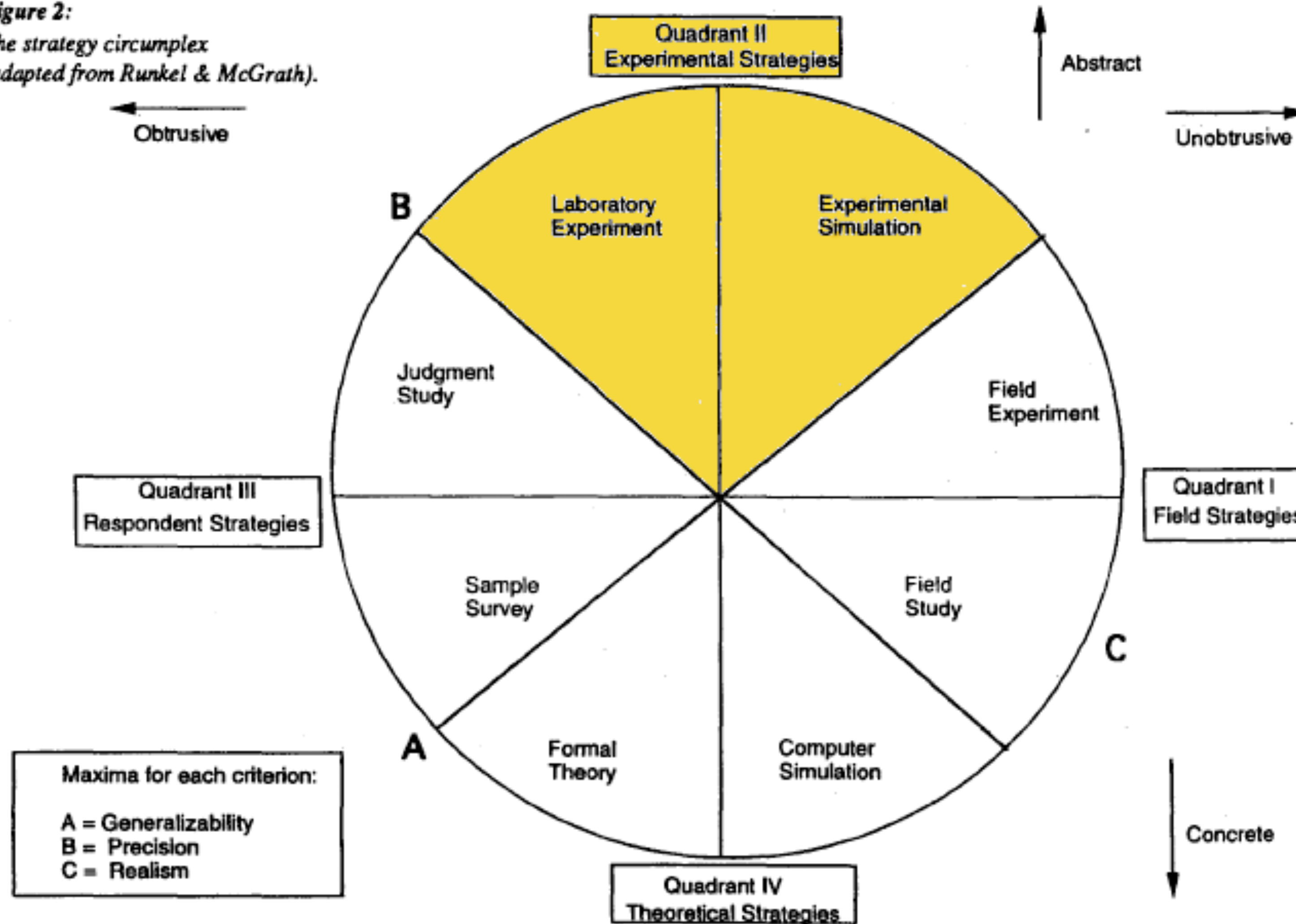




# Taxonomy of Methods

[McGrath et al. 1994]

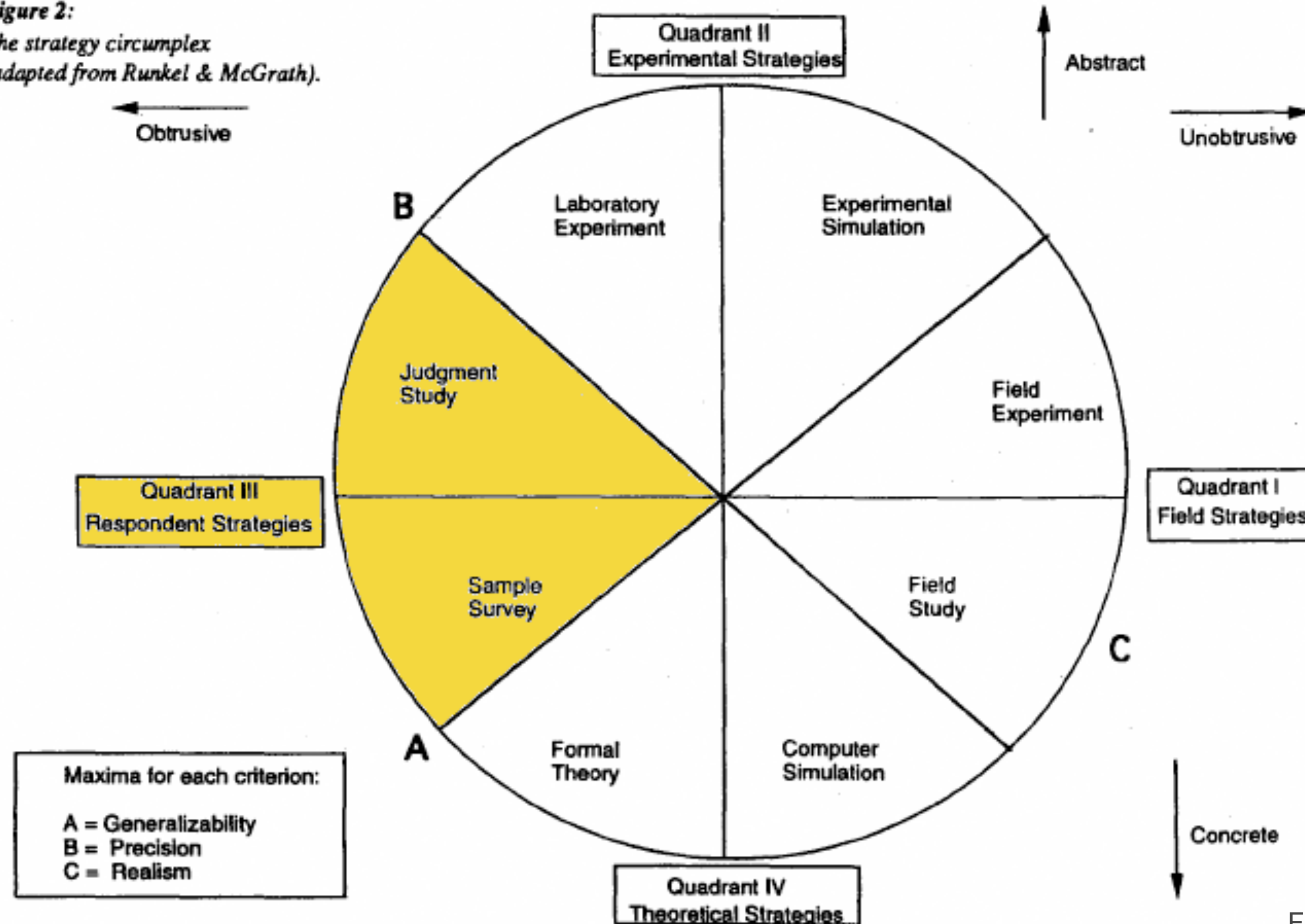
Figure 2:  
The strategy circumplex  
(adapted from Runkel & McGrath).



# Taxonomy of Methods

[McGrath et al. 1994]

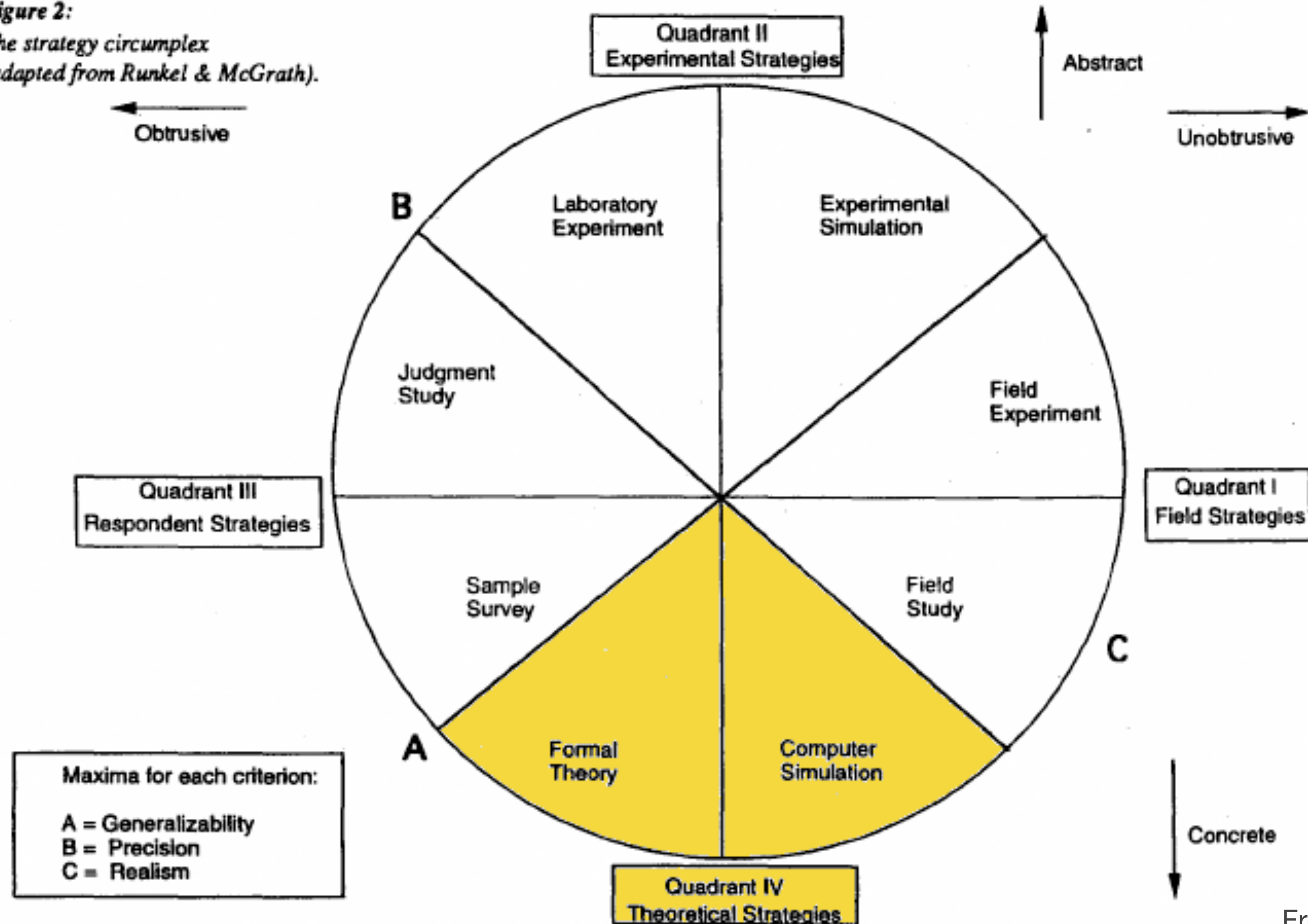
Figure 2:  
The strategy circumplex  
(adapted from Runkel & McGrath).



# Taxonomy of Methods

[McGrath et al. 1994]

Figure 2:  
The strategy circumplex  
(adapted from Runkel & McGrath).



# Évaluation et tests

- Introduction
- Approches d'évaluation
- **Méthodes analytiques**
- Méthodes empiriques



# Les grands types d'évaluation analytique

## 1. Basée sur des modèles

- Évaluation selon des modèles d'interaction
- Voir [ce cours](#) de Gilles Bailly

## 2. Basée sur l'inspection

- Review d'experts / critique du design
- Cognitive walkthrough
- Évaluation heuristique

# Évaluation basée sur des modèles

## GOMS (Goals, Operators, Methods, and Selection rules)

- Les objectifs (Goals) correspondent à ce que l'utilisateur essaie de faire
- Les opérateurs sont les actions faites pour atteindre cet objectif.
- Les méthodes sont les séquences d'opérateur qui permettent d'accomplir un objectif. Il peut y avoir plus d'une méthode pour un même objectif, dans ce cas :
- Les règles de sélection (Selection rules) sont utilisées pour décrire quand un utilisateur voudrait choisir tel méthode plutôt qu'une autre. On ignore souvent ces règles dans une analyse GOMS simple.

## KLM

- On analyse une action et on la découpe en étapes atomiques
- On cherche la durée de chaque étape dans une table
- On prédit la durée de l'action complète
- *Permet la prédire avant d'implémenter !*

# GOMS analysis

```

GOAL: EDIT-MANUSCRIPT
.   GOAL: EDIT-UNIT-TASK ... repeat until no more unit tasks
.   .   GOAL: ACQUIRE UNIT-TASK
.   .   .   GOAL: GET-NEXT-PAGE ... if at end of manuscript page
.   .   .   GOAL: GET-FROM-MANUSCRIPT
.   .   .   GOAL: EXECUTE-UNIT-TASK ... if a unit task was found
.   .   .   GOAL: MODIFY-TEXT
.   .   .   .   [select: GOAL: MOVE-TEXT* ...if text is to be moved
.   .   .   .   .   GOAL: DELETE-PHRASE ...if a phrase is to be deleted
.   .   .   .   .   GOAL: INSERT-WORD] ... if a word is to be inserted
.   .   .   .   VERIFY-EDIT

```

\*Expansion of MOVE-TEXT goal

```

GOAL: MOVE-TEXT
.   GOAL: CUT-TEXT
.   .   GOAL: HIGHLIGHT-TEXT
.   .   .   [select**: GOAL: HIGHLIGHT-WORD
.   .   .   .   MOVE-CURSOR-TO-WORD
.   .   .   .   DOUBLE-CLICK-MOUSE-BUTTON
.   .   .   .   VERIFY-HIGHLIGHT
.   .   .   .   GOAL: HIGHLIGHT-ARBITRARY-TEXT
.   .   .   .   .   MOVE-CURSOR-TO-BEGINNING           1.10
.   .   .   .   .   CLICK-MOUSE-BUTTON             0.20
.   .   .   .   .   MOVE-CURSOR-TO-END             1.10
.   .   .   .   .   SHIFT-CLICK-MOUSE-BUTTON        0.48
.   .   .   .   .   VERIFY-HIGHLIGHT]           1.35
.   .   .   GOAL: ISSUE-CUT-COMMAND
.   .   .   .   MOVE-CURSOR-TO-EDIT-MENU           1.10
.   .   .   .   PRESS-MOUSE-BUTTON              0.10
.   .   .   .   MOVE-CURSOR-TO-CUT-ITEM         1.10
.   .   .   .   VERIFY-HIGHLIGHT              1.35
.   .   .   .   RELEASE-MOUSE-BUTTON          0.10

```

...

# KLM

Description	Operation	Time (sec)
Reach for mouse	H[mouse]	0.40
Move pointer to "Replace" button	P[menu item]	1.10
Click on "Replace" command	K[mouse]	0.20
Home on keyboard	H[keyboard]	0.40
Specify word to be replaced	M4K[word]	2.15
Reach for mouse	H[mouse]	0.40
Point to correct field	P[field]	1.10
Click on field	K[mouse]	0.20
Home on keyboard	H[keyboard]	0.40
Type new word	M4K[word]	2.15
Reach for mouse	H[mouse]	0.40
Move pointer on Replace-all	P[replace-all]	1.10
Click on field	K[mouse]	0.20
Total		10.2

# Limites

Des prédictions valides pour un utilisateur expert qui ne fait pas d'erreur

- les experts font aussi des erreurs !
- pas de prise en compte des utilisateurs novices ou intermédiaires qui font des erreurs occasionnelles.
- il existe des extensions qui essaient de modéliser l'apprentissage

Toutes les tâches ont un objectif clair

- Beaucoup de tâches ne sont pas si dirigées spécialement en design UX.

Ne prend pas en compte les différences individuelles entre utilisateurs

- Basé sur des moyennes statistiques

Ne prend pas en compte les aspect sociaux et organisationnels du produit

Pas d'information sur la qualité d'utilisation et le plaisir provoqué par le produit.

Pas représentatif des théories actuelles sur la cognition humaine

- Supposition d'un model linéaire de la cognition avec une activité faite à la fois.

# Les grands types d'évaluation analytique

## 1. Basée sur des modèles

- Évaluation selon des modèles d'interaction
- Voir le cours de Gilles Bailly

## 2. Basée sur l'inspection

- Review d'experts / critique du design
- Cognitive walkthrough
- Évaluation heuristique
- Voir le cours d'IHM de L3



# Inspections et critiques d'experts

Tout au long du processus de développement

Conduite par des développeurs et des experts (internes ou externes)

Outil pour identifier des problèmes

Peut aller d'une heure à une semaine de travail

Préférer une approche structurée

- Les reviewers doivent pouvoir communiquer sur tous les problèmes (sans fâcher l'équipe)
- Les critiques ne doivent pas être aggressive envers les développeurs / designers
- L'objectif principal est d'identifier les problèmes (pas leur source)

Des solutions peuvent être suggérées a l'équipe

# Les critères ergonomiques

## usability guidelines

- Don Norman's principles:
  - visibility, affordances, natural mapping, and feedback
- Ben Shneiderman's 8 Golden Rules of UI design
- Bruce Tognazzini's 16 principles:
  - <http://www.asktog.com/basics/firstPrinciples.html>
- Christian Bastien's Ergonomic Criteria
- Jakob Nielsen's Heuristics

# +/- of expert evaluation

Results of informal reviews and inspections are often directly used to change the product

- ... still state of the art in many companies
- The personal view of the CEO, or his buddies ...

Really helpful evaluation

- Is explicit
- Has clearly documented findings
- Can increase the quality significantly

Expert reviews and inspections are a starting point for change

# Méthodes d'évaluation

- Introduction
  - ~~Approches d'évaluation~~
  - ~~Méthodes analytiques~~
  - Méthodes empiriques
- **Concevoir une expérience**
  - Exemples
  - Les bases
  - La structure d'une expérience
  - Mener une expérience
  - Collecter les données
- Analyser les données